The Geneva Papers on Risk and Insurance - Issues and Practice
https://doi.org/10.1057/s41288-019-00143-9

Contract elements, growing conditions, and anomalous claims behaviour in U.S. crop insurance

Sungkwol Park[1] · Barry K. Goodwin[2] · Xiaoyong Zheng[2] · Roderick M. Rejesus[2]

Received: 7 January 2019 / Accepted: 17 July 2019

Abstract

We investigate contract elements and growing conditions associated with anomalous claims behaviour in the U.S. Federal Crop Insurance Program. In this study the measure of "anomalous claims behaviour" is based on the number of producers (in a county) placed on the "Spot Check List" (SCL)—a list generated from government compliance efforts that aim to detect and deter fraud, waste, and abuse in the U.S. Federal Crop Insurance Program. Using county-level data and various econometric approaches that control for features of this data set (e.g., the count nature of the dependent variable, censoring, potential endogeneity, and spatial/temporal dependence), we find that the following crop insurance contract attributes influence the extent of anomalous claims behaviour in a county: (a) the ability to insure individual fields through "optional units"; (b) the coverage level choice; and (c) the total number of acres insured. In addition, our empirical analyses suggest that anomalous claims behaviour significantly increases when extreme weather events occur (e.g., droughts, floods) and when economic conditions are unfavourable (i.e., high input costs that lower profit levels). Results from this study have important implications for addressing potential underwriting vulnerabilities in crop insurance contracts and the frequency of more rigorous compliance inspections.

Keywords Spot Check List · Insurance fraud · Crop insurance · Simulated maximum likelihood estimation · Control function approach · Block bootstrap

---

Electronic supplementary material The online version of this article (https://doi.org/10.1057/s41288-019-00143-9) contains supplementary material, which is available to authorised users.

✉ Roderick M. Rejesus
rmrejesu@ncsu.edu

[1] Ministry of Economy and Finance, Seoul, South Korea

[2] Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695, USA

Published online: 26 August 2019

Introduction

The Risk Management Agency (RMA) of the U.S. Department of Agriculture (USDA) is in charge of administering the U.S. Crop Insurance Program. As part of its efforts to detect and deter fraud, waste, and abuse in the Crop Insurance Program, the RMA developed and implemented the so-called Spot Check List (SCL) programme in 2001 (USDA-RMA 2011). Under this SCL effort, the RMA and their partners utilise complex and proprietary algorithms to analyse a massive data warehouse that contains extensive crop insurance contract data as well as information from other related databases (e.g., weather data and other administrative data from other USDA agencies). The aim of the SCL process is to detect individual producers whose claims behaviour demonstrates atypical patterns that are indicative of potential fraud, waste, and abuse (or generally, moral hazard).[1]

One of the main outputs from this process is the SCL itself—an annual list of insured farmers who are identified based on objective and data-driven statistical techniques and whose loss experiences are considered “anomalous” relative to similar producers in the same geographic area (i.e., typically within a county) producing the same crop and using the same cropping practices. Therefore, the number of SCL producers in a county can be regarded as a measure of “anomalous claims behaviour” suggestive of the extent of crop insurance fraud, waste, and abuse in that county.[2] Everything else being equal, larger numbers of SCL producers (in a county) potentially indicate that there are likely to be more fraudulent claims or that there is more claims misrepresentation in that county.

The objective of this study is to examine crop insurance contract elements and growing conditions that likely influence the degree of anomalous claims behaviour in a county. We first examine whether the underwriting design of the crop insurance contract itself is a contributing factor, since opportunistic abuses and misrepresentations can be prompted by flexible insurance options and provisions. We focus on the unit structure, coverage level, and type of insurance chosen by the insured farmers. The unit structure defines how acres are insured.[3] Indeed, Knight and Coble

[1] Although moral hazard is the general term used in insurance economics to represent fraud, waste, or abuse, we use the combined term “fraud, waste, and abuse” throughout this paper because these are the exact terms used in RMA documents related to the SCL process. Studies that have explored the extent of moral hazard in crop insurance include Coble et al. (1997) and Roberts et al. (2006).

[2] As mentioned in USDA-RMA (2018), being included on the SCL does not indicate that a producer has explicitly engaged in fraud. Rather, it implies that the claims behaviour of a producer on the SCL is not consistent with that of other similar producers in the same geographic area, and therefore warrants more extensive investigation. In addition, we should note here that one drawback of using deviations from “peers” as a measure of anomalous behaviour is that its likelihood of accurately detecting fraud would likely be lower when all (or a majority of) farmers in the area are also engaging in potential fraud, waste, and abuse (i.e., there can be no deviation from peers because everybody is doing the same thing). Although this situation is certainly possible, anecdotal evidence and conversations with RMA compliance personnel suggest that this is unlikely.

[3] Under the current Federal Crop Insurance Program, producers have four options: optional, basic, enterprise, and whole-farm units. Under the optional unit, each field and crop can be insured separately while under other options, several units are combined. Specifically, basic units combine all of the owned and cash-rented acres in the same county by the same producer, but share-leased units or units of a different crop cannot be combined. Enterprise units combine all acreage of the same crop by the insured in

(1999) found that the loss cost ratios[4] tend to be higher for optional units than for basic units. Considering that producers can choose their coverage level between 50 and 85% (in 5% increments) under the buy-up coverage option in U.S. crop insurance, we also examine whether higher coverage levels lead to more anomalous claims behaviour.[5] Walters et al. (2015) found evidence that producers obtained excess returns by selecting optional units and buy-up coverage. Lastly, producers can choose between yield-based and revenue-based policies. Revenue-based policies insure against losses due to both low crop yields and price declines, while yield-based policies offer protection against low yields only. We consider whether less exposure to price risk under the revenue-based policies contributes to a decrease in anomalous claims behaviour. It is also possible that revenue-based policies lead to more anomalous claims behaviour, as revenue-based policies are more costly than yield-based ones.

In addition to crop insurance contract characteristics, we also examine the roles of two additional categories of variables in determining anomalous claims behaviour. First, unfavourable economic conditions in a particular year may be another motivating factor for producers to commit insurance claims fraud, waste, and abuse. For this, we use a county-level net income variable (the difference between income and expenses) as a measure of producers' profits (and the general economic environment). Also, anomalous claims behaviour may be triggered by extreme weather conditions. Given the fact that not all claims are audited, when severe drought or extremely wet conditions (e.g. floods) occur, producers may be more likely to engage in fraudulent activities because the extreme weather conditions make it more difficult to distinguish between fraudulent claims and legitimate ones. For this reason, we include a rich set of weather variables in our empirical analysis.

Empirically estimating the effects of crop insurance contract elements and growing conditions on the extent of county-level anomalous claims behaviour is challenging for a number of reasons. First, the data available to us has three important features (and limitations) that need to be taken into account in the estimation: (1) our data set is at the county-level with a panel (or longitudinal) structure; (2) our dependent variable, the number of SCL producers in a county, is an integer count variable; and (3) our dependent variable is left-censored due to government regulations on data confidentiality when reporting SCL data.[6] To accommodate the three data features described, we estimate a random effects count model with censoring,

Footnote 3 (continued)

a county into one unit. In other words, share-leased acres can be combined with owned and cash-rented acres in one enterprise unit. Whole-farm units combine all crops by the insured in the same county. All four unit structures are illustrated in section "Unit structure" with an example.

[4] Loss cost ratio is the ratio of indemnities to liabilities.

[5] Producers can purchase the minimum catastrophic coverage (CAT) that will protect up to 50% of their expected yield/revenue. Producers can purchase (buy-up) higher levels of coverage with the option to insure up to 85% of the expected yield/revenue.

[6] The SCL data are highly confidential and were only available at a county aggregate level and only reported to us if the number of SCL cases was 4 or more, making the SCL count censored.

using the simulated maximum likelihood approach.[7] Another important feature of the data set (and the variables included in the specification) is that several explanatory variables in our regression models are potentially endogenous. Therefore, we also employ a two-stage least squares (2SLS) and a control function approach (Hausman 1978; Wooldridge 1997, 2014; Terza et al. 2008) to correct for possible bias due to endogeneity.

Our results provide strong evidence that county-level variables associated with characteristics of the crop insurance contracts purchased are strong predictors of anomalous claims behaviour. For instance, a larger proportion of acres insured as optional units is associated with more anomalous claims behaviour. Higher average coverage levels (in a county) and a larger number of insured acres are also strongly linked with increased incidence of anomalous claims behaviour. In addition, our empirical analyses also indicate that unfavourable economic and weather conditions strongly influence the extent of anomalous claims behaviour in a county.

The crop insurance contract elements associated with anomalous claims behaviour are potential underwriting vulnerabilities that RMA needs to address to further discourage fraud, waste, and abuse in the system. Further study is required to determine the appropriate underwriting or premium-rating adjustments needed to curb the anomalous claims behaviour linked to these "vulnerable" contract elements. Identifying these vulnerabilities can also help provide directions as to how RMA and the AIPs can improve and better utilise their fraud detection algorithms (USDA Office of the Inspector General (OIG) 2017). On the other hand, our empirical results regarding insureds potentially taking advantage of unfavourable economic and weather conditions to engage in fraud, waste, or abuse indicate that fraud auditing standards and compliance inspection efforts may need to be tightened in times when adverse economic and weather conditions are affecting the agricultural economy. Moreover, given that the algorithms used by RMA to detect anomalous behaviour are confidential,[8] this study also contributes to the literature by providing the risk management research community with a third-party view of likely factors considered in the SCL detection process.

The remainder of this paper is organised as follows. The next section introduces the SCL programme and describes the unit structure in U.S. federal crop insurance. In the third section, we describe the data and variables used in our empirical analyses. In the fourth section, we discuss the econometric model utilised to account for the special features of our data. In the fifth section, the main findings from our estimations are discussed. In the sixth section, we conduct robustness checks to examine the sensitivity of our results when using alternative clustering procedures and specifications. Concluding comments and policy implications are provided in the final section.

[7] Although this is our "preferred" estimation procedure (since it accounts for all three data features above), we also utilised fixed effects linear models and fixed effects count models that account for some (but not all) of the data features described above.

[8] The algorithms and techniques used to determine individuals to be included in the SCL are purposely confidential (and not publicly available). RMA does not want insureds to take advantage of this type of information to alter their behaviour and avoid detection.

Background

Objective and data-driven statistical techniques are employed to annually develop a list of SCL producers whose loss experience is evaluated as "anomalous" relative to similarly situated producers in the same geographic area (typically within the county) producing the same crop and using the same cropping practices. To develop the SCL, previous research and "on-the-ground" observations gathered by RMA field staff and partner insurance companies [called Approved Insurance Providers (AIPs)] are first utilised to identify different "scenarios" that suggest potential fraud, waste, or abuse. For example, one scenario-based detection algorithm may pertain to finding those producers who have large multi-year losses that are consistently higher than their peers in the same county. Another algorithm may aim to detect behaviour consistent with known fraud schemes that have previously been recognised.

All producers flagged by these detection procedures are then used to create a pool of producers to be included in the SCL. The SCL developed for producers of spring-planted crops (e.g. corn, soybeans) that are based on data analysed through a particular crop year (say, in 2017, where data through December 2017 are analysed) is typically finalised no later than the first quarter of the following year (i.e., no later than 1 April 2018). The finalised SCL is then forwarded to the local USDA Farm Service Agency (FSA) county offices where the SCL farmers are located and to the AIPs whose clients are included in the list. Soon after, the AIPs send a formal letter to their clients on the list, informing them of their inclusion and that their operations are subject to inspection and/or review during the growing season. The FSA county offices and/or the AIPs then conduct infield inspections and/or policy reviews on the SCL producers. In practice, due to time and resource constraints, only a subset of SCL farmers are inspected or reviewed.[9]

Unit structure

Unit structure is one of the main contract elements we examine in this study. The options are complex and merit additional discussion. Consider a producer who has cropland units in a county shown in Fig. 1.[10] The producer has six units, identified by the letters A through F. The fields for corn are from A to C and those for soybeans are from D to F. In addition, since unit choices are based on ownership

[9] Note that from 2001 to 2011 the FSA had sole responsibility for conducting infield inspections of all SCL producers (i.e., both growing season and pre-harvest inspections) to assess whether the condition of the insured crop is consistent with other non-SCL producers in the area. Beginning in 2012, AIPs assumed responsibility for inspecting a subset of producers on the SCL, with the FSA still responsible for the remaining producers. The AIP inspection is more comprehensive than that of the FSA in the sense that they perform both infield inspections and a full policy review. For a more detailed description of what is involved in an AIP inspection and full policy review, see: https://legacy.rma.usda.gov/pubs/ra/sraarchives/19sra.pdf.

[10] An optional unit is formally defined as all land under a common ownership structure in a section. This often represents a single field or a collection of smaller fields.

structure as well as the geographic location, we need to know whether each field is owned, cash-rented, or share-rented by the producer.

As an example, let us assume that the producer owns fields A and D, rents fields B and E on a cash basis, and has a crop share arrangement on fields C and F. The producer could choose to insure the fields with six optional units if field-specific production records were available and the boundaries of the units were readily discernible (i.e., three for corn and three for soybeans). An insurance guarantee is assigned to each optional unit, and each unit would stand alone when determining indemnities and would be charged a higher premium rate in recognition of the higher risk.

The farmer could also choose to insure the fields with three other unit structures. A basic unit[11] consists of all croplands of a single crop that are either owned or cash-rented. In the case of a single crop share arrangement, a separate basic unit is available. Thus the producer in our example can have two basic units of corn and two basic units of soybeans (i.e., a total of four basic units). The first corn basic unit consists of fields A and B. The other basic unit consists of field C. The first soybean basic unit consists of fields D and E. The second basic unit consists of field F. Each basic unit has its own insurance guarantee. The move to basic units from optional units may decrease the frequency with which indemnities are received. Because of this potential decrease in indemnities, insurance premiums are lower for basic units than for optional units.[12]

An enterprise unit consists of all acreage of a crop in the county. Therefore, our producer could form two enterprise units: one for corn and one for soybeans. The expected indemnities (and insurance premiums) are lower under enterprise units than under basic units because the losses are aggregated across the unit.

A whole-farm unit consists of all croplands in the same county placed in a single insurance unit. Thus the farm in our example could insure all acres of cropland together, that is, the corn and soybean fields would be insured together in one whole-farm unit.

Note that the government's premium subsidy also varies by unit type. In general, the premium subsidies for basic and optional units are lower than those for enterprise and whole-farm units. Moreover, whole-farm units have slightly higher premium subsidies than enterprise units, especially at higher coverage levels (75% and above). Therefore, the unit type eventually chosen by farmers to be insured is influenced by the amount of premium subsidy for each unit type (at a particular coverage level).[13]

[11] The insured automatically qualifies for basic units without exception (USDA-RMA 2015 Crop Insurance Handbook).

[12] Since 1988, producers have received a fixed 10% discount on their premiums if they do not insure optional units.

[13] In the next section, we discuss the general pattern of unit choices over time that was largely affected by changes in premium subsidies. Current premium subsidy levels by unit type and coverage level can be seen in Motamed et al. (2018, p. 11). Moreover, see Babcock and Hart (2005) for more detailed explanations about unit structures, including a calculation of liabilities and indemnities for each unit choice.

Data

We obtained county-level "Spot Check List" (SCL) data from USDA-RMA over the 2001–2015 time period through a Freedom of Information Act (FOIA) request. Because data on county-level unit structure—a key contract element examined in this study—are only available from 2002 (USDA-RMA, 2019),[14] our regression analysis only utilises data from 2002 to 2015. The SCL data cover insured producers of four major row crops (corn, soybeans, wheat, and cotton) in addition to tobacco. Also, only yield-based and revenue-based individual policies are considered.[15] As a result, our data account for 69.59% of all crop insurance policies for which acreage had been reported to USDA-RMA between 2002 and 2015. These data come from 2162 counties across all U.S. states, except Alaska, Hawaii, and Rhode Island.[16]

More importantly, due to government regulations regarding data confidentiality, the number of SCL producers in a county is only reported in our data set if the county had at least four producers on the SCL in that year. We therefore cannot exactly identify the number of producers on the SCL in a county for a particular year when the number of producers on the SCL was less than four. Our empirical specification below is designed to accommodate this important data feature. The number of counties with more than three SCL producers from 2001 to 2015 are presented in Table 1. The numbers ranged from 124 to 308 during the sample period. On average, 235 counties had at least four SCL producers in a particular year and these counties had approximately seven SCL producers every year. Table 2 summarises the frequency distribution of counties with each number of SCL producers by year. Figure 2 provides the spatial distribution of the total number of SCL producers from 2001 to 2015.[17] We note that counties with substantial SCL producers are scattered throughout the continental U.S., with some clustering in the upper Midwest, the Dakotas, the Plains (i.e., Kansas, Nebraska and the Texas Panhandle), and the southeastern states (i.e., North Carolina, South Carolina, Georgia, and Florida).

Figures A.1 and A.2 in the Online Appendix provide the spatial distributions of the number of SCL producers for selected years between 2001 and 2015. When the SCL programme started in 2001, the Dakotas and the Plains had larger numbers of SCL producers, but the clusterings gradually disappeared during the 10-year period from 2001 to 2010 (as seen in Online Appendix Fig. A.1).

Figure A.2 in the Online Appendix presents the spatial distributions of the number of SCL producers from 2012 to 2015. It appears that for these 4 years, a number of SCL producers were in Iowa, Missouri, Illinois, and Kansas.[18]

[14] See: https://www.rma.usda.gov/data/sob/scc/index.html.

[15] Other "less popular" plans like the Area Risk Protection Insurance (ARPI) and Whole Farm Revenue Protection (WFRP) Insurance policies are not considered in this study.

[16] Crop insurance policies from 2002 to 2015 were sold in 2831 counties across all 50 U.S. states.

[17] Because of the limitation on the number of SCL producers reported in our data, if the number of SCL producers in a year was less than four for a county, then the number of SCL producers in this figure was coded as zero.

[18] It is important to emphasise here that the SCL procedure is national in scope. Even with these geographical SCL "clusterings" observed over time, there was no explicit attempt to "target" a specific

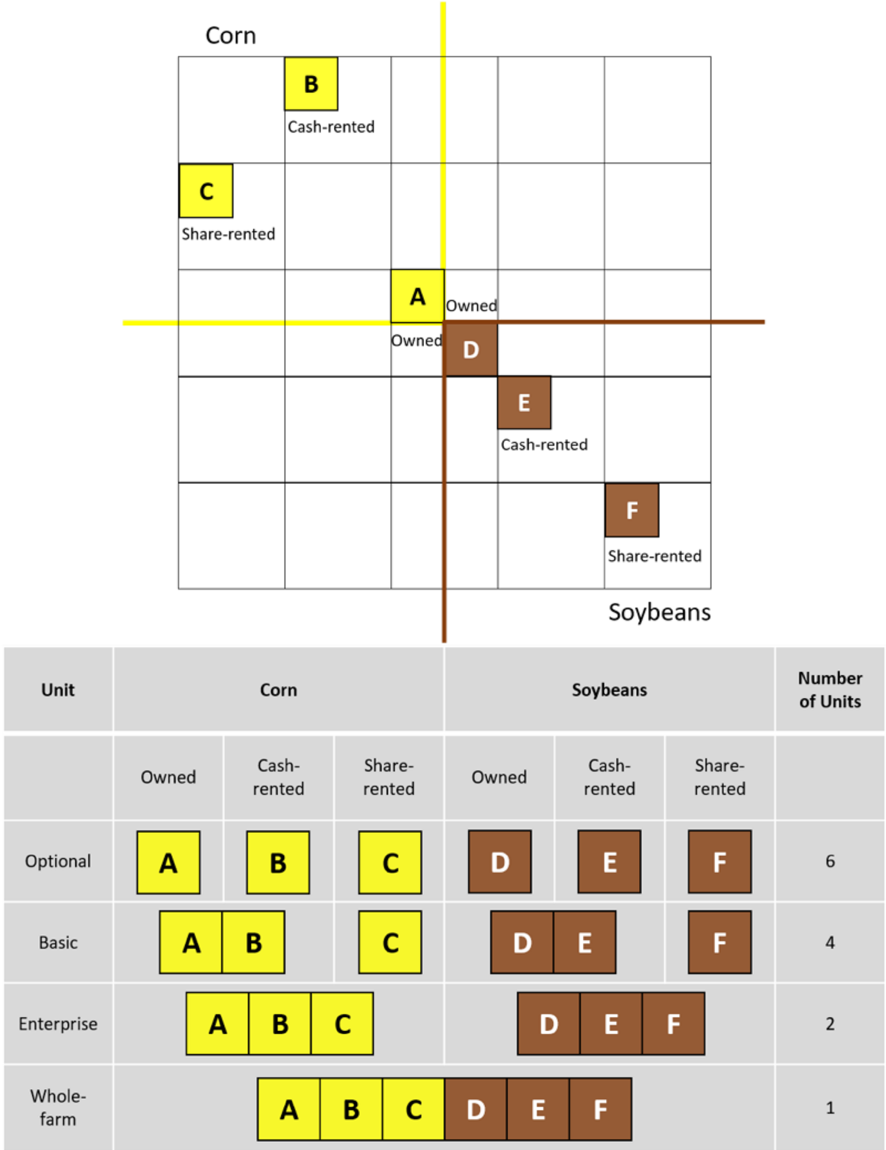

Fig. 1 Example of unit structure

We constructed a number of county-level variables to represent contract elements and growing conditions that could potentially influence anomalous claims behaviour. First, for factors related to crop insurance contract design, we use the county-level

Footnote 18 (continued)
region in the U.S. or a particular set of crops. These SCL "clusterings" over time are simply a result of objective and data-driven algorithms applied nationally in order to detect anomalous behaviour.

Table 1 Number of counties with more than 3 SCL producers

Year	Number of counties	Percent (%)	Mean	Std. Dev.	Max
2001	210	9.57	10.71	12.07	98
2002	205	9.34	6.63	2.86	17
2003	288	13.13	7.31	3.91	26
2004	218	9.94	7.28	4.27	38
2005	178	8.11	6.43	3.52	23
2006	124	5.65	7.01	3.80	27
2007	265	12.08	6.24	2.49	17
2008	289	13.17	6.31	3.17	30
2009	228	10.39	6.01	2.15	14
2010	191	8.71	6.06	2.40	18
2011	239	10.89	5.69	2.13	14
2012	236	10.76	5.55	1.95	13
2013	306	13.95	6.33	2.23	14
2014	308	14.04	6.21	2.18	11
2015	233	10.62	5.50	1.78	12

SCL data include 2194 counties each year

Table 2 Frequency distribution of counties with different numbers of SCL producers

Year	Number of SCL producers										
	≤3	4	5	6	7	8	9	10	11–20	21–30	> 30
2001	1984	55	28	23	12	19	7	9	37	6	14
2002	1989	56	38	29	21	18	12	14	17	0	0
2003	1906	78	45	39	30	22	9	14	48	3	0
2004	1976	63	39	25	21	11	15	6	36	1	1
2005	2016	59	45	19	15	12	5	5	16	2	0
2006	2070	35	27	14	10	9	4	7	17	1	0
2007	1929	77	66	34	18	20	14	23	13	0	0
2008	1905	96	52	46	29	17	18	15	14	2	0
2009	1966	71	49	38	20	14	15	13	8	0	0
2010	2003	60	46	30	13	11	7	13	11	0	0
2011	1955	91	56	35	20	12	4	9	12	0	0
2012	1958	96	50	37	23	8	6	5	11	0	0
2013	1888	83	61	40	34	26	23	28	11	0	0
2014	1886	98	57	33	24	36	23	31	6	0	0
2015	1961	92	53	36	19	11	12	7	3	0	0

crop insurance experience data publicly available from USDA-RMA (USDA-RMA, 2019). As discussed in the introductory section, we focus on unit structure, coverage level, and insurance type. To examine whether the ability to separately insure individual fields through optional units influences the potential risk of fraud, waste, and abuse, we use the county-level percentage of acres insured as optional units.

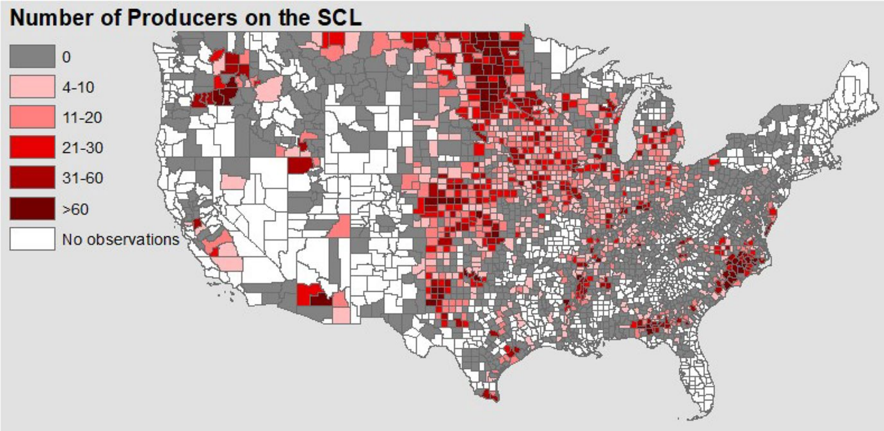

Fig. 2 Spatial distribution of total number of SCL producers from 2001 to 2015

Figure 3 shows the percentages of acres insured under the four different types of units from 2002 to 2015, and Fig. A.3 in the Online Appendix presents the percentages of acres insured as optional units by crop over time. In 2002, on average, 57% of the acres insured were contracted as optional units, and this percentage increased to 62% in 2008. However, it decreased considerably between 2008 and 2015, partly because the government raised the premium subsidies for the use of enterprise units by a significant amount in 2009 (the effect can be clearly seen in Fig. 3), and in 2015, only 32% of the acres were contracted as optional units. Figure A.4 in the Online Appendix presents the spatial distribution of the percentage of acres insured as optional units from 2002 to 2015. We note that counties with a higher percentage of optional units (over 60%) were concentrated in the Dakotas, the Plains (i.e., Kansas, Nebraska, and the Texas Panhandle), and a few southeastern states (i.e., North Carolina, South Carolina, and Georgia).

In addition to the unit structure, we also examine coverage level as a potential factor influencing anomalous claims behaviour. We created a county-level acreage-weighted average coverage level variable. Figure 4 shows that average coverage levels for all crops together and individually increased gradually over the sample period from about 67% for all crops in 2002 to about 74% in 2015. Figure A.5 in the Online Appendix presents the spatial distribution of the average coverage levels over the sample period. As can be seen, average coverage levels were not evenly distributed, and particularly the Corn Belt states in the Midwest had much higher coverage levels than other regions.[19] The third insurance contract variable we examine is the insurance type. Figure 5 and Figure A.6 in The Online Appendix present the trend and spatial distribution of the percentage of acres insured under revenue-based policies. It is clear that during the sample period the percentage of acres insured under revenue-based policies increased gradually from about 52% in 2002 to about 90%

[19] For further discussion on the coverage level, see Schnitkey and Sherrick (2014).

in 2015 for all crops combined. The same pattern holds for individual crops as well. Finally, we include the number of acres insured in the county as an explanatory variable. Larger counties tend to have more producers and this may lead to more producers on the SCL. This information is also publicly available from USDA-RMA.

For the measure of economic conditions, we collected county-level income data on producers' cash receipts on crops from the U.S. Bureau of Economic Analysis (BEA) (U.S. Department of Commerce-BEA, 2019).[20] We also considered the county-level expenditure data by adding up production expenses for seeds, fertiliser, chemicals, petroleum products (fuel), hired labour, and all other expenses, which are available from the same source as the income data. We then computed net income as the difference between income and expenditure.

Lastly, we created a rich set of weather variables to be included in the specification (and to examine whether adverse weather conditions affect anomalous claims behaviour). First, we created several yearly weather variables that represent the occurrence of extreme weather conditions (i.e., the average of monthly measures of extreme heat, extreme drought, and extreme wet conditions). For the measure of extreme heat conditions, we use county-level total degree days above 30 °C during the months of July and August using the method of Schlenker and Roberts (2006).[21] For extreme drought and extreme wet conditions (e.g. flood), we collected the monthly Palmer Drought Severity Index[22] from April to October at the state level from the National Oceanic and Atmospheric Administration (NOAA) and constructed two variables, one for dryness (negative values) and the other for wetness (positive values) (NOAA, 2019). In addition, we also collected monthly county-level data on average precipitation (in 10 cm), minimum temperature (°C), and maximum temperature (°C) for the growing months (from April to October), based on the work of Schlenker and Roberts (2009) and available data from PRISM.[23] Table 3 lists all the variables used in our estimation and the corresponding data sources. Summary statistics are displayed in Table 4.

Estimation

There are three main features of our data set. First, it is a panel data set at the county level. Second, the dependent variable—the number of producers on the SCL in each county—is a count variable. Third, the dependent variable is left-censored, that is, it takes the value of zero when the number of producers on the SCL is less than 4 (i.e., we only know that the number can be 0, 1, 2, or

[20] See: https://www.bea.gov/regional/.

[21] For a detailed derivation and further discussion, see Schlenker and Roberts (2006) and the SI Appendix of Schlenker and Roberts (2009).

[22] To identify the abnormality of a drought in a region for a particular month, short-term drought index (i.e., the Palmer Z index) is used. It indicates that the Z Index represents monthly drought conditions with no memory to previous monthly moisture deficits or surpluses.

[23] See http://www.prism.oregonstate.edu. Further note that the "average" values here were averaged across the days in a month and across different weather stations in a county.

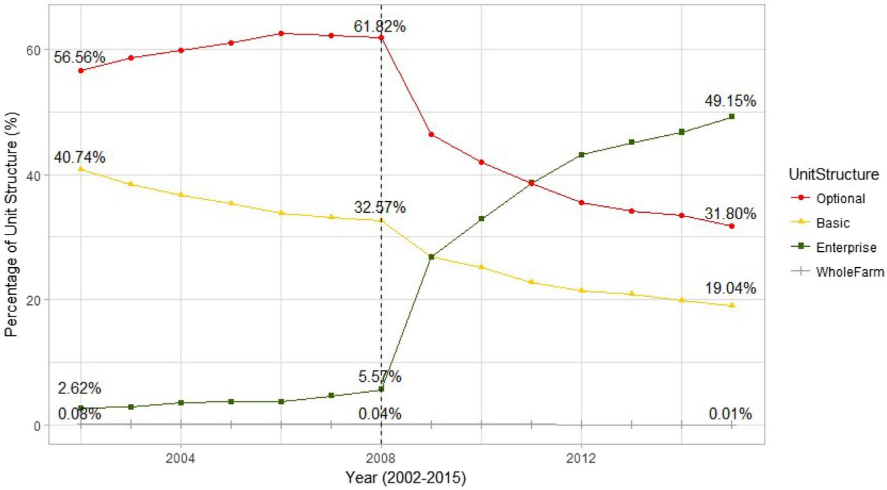

Fig. 3 Percentages of acres insured as different types of units from 2002 to 2015

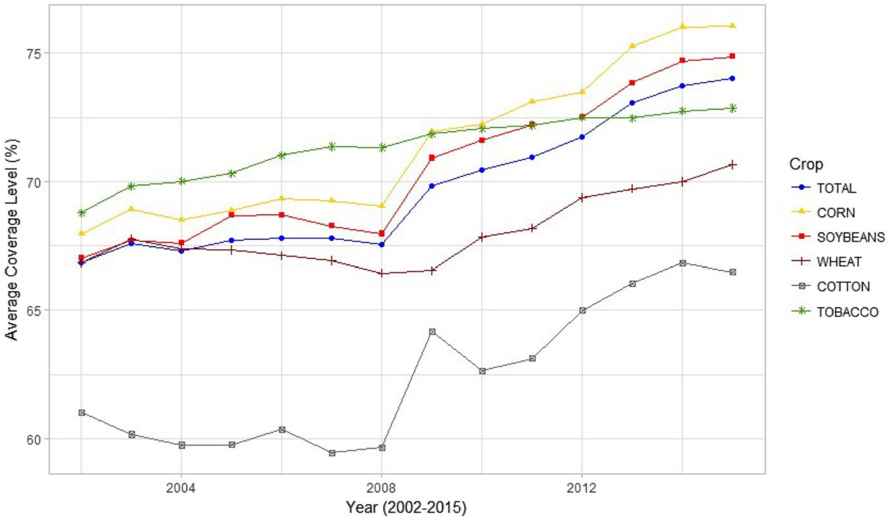

Fig. 4 Average coverage levels from 2002 to 2015

3). We estimate three empirical models. Each model has its own advantages and disadvantages.

Formally, let y_{it} denote the number of producers on the SCL in county i and year t (i.e., SCL_{it}). First, we estimate the following fixed effects linear model,

$$y_{it} = X'_{it-1}\Theta + \alpha_i + \epsilon_{it}, \tag{1}$$

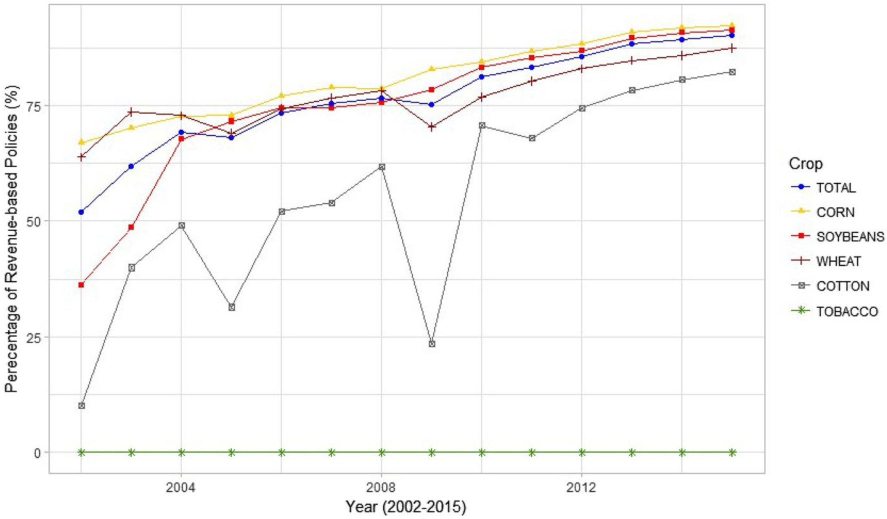

Fig. 5 Percentages of acres insured under revenue-based policies from 2002 to 2015

Table 3 Variable definitions and sources

Variables	Description and sources
1. Dependent variable	
SCLProducerCount	County-level number of producers in Spot Check List
2. Explanatory variables	
Insurance data[a]	
OptionalUnit	Percentage of acres insured as optional units
CoverageLevel	Average coverage level weighted by acres insured
RevenueInsurance	Percentage of acres insured under revenue-based policies
AcresInsured	Number of acres insured (in 1000 acres)
Profit data[b]	
NetIncome	The difference between cash receipts from crops and total expenditure on seeds, fertiliser, chemicals, petroleum products, labour, and all other expenses (in USD 100 million)
Yearly weather data	
Dday30C[c]	Total degree days above 30 °C (Celsius and days), Jul–Aug
Drought[d]	Palmer Z index for drought level, Apr–Oct
Wetness[d]	Palmer Z index for wetness level, Apr–Oct
Other monthly weather data	
Prec[c]	Precipitation (in 10 cm), Apr–Oct
tMin, tMax[c]	Averages of Min. (Max.) temperatures (Celsius), Apr–Oct

[a]Reproduced from Summary of Business of RMA (County level)

[b]BEA (Bureau of Economic Analysis): CA45 Farm income and expenses (County level)

[c]Reproduced based on Schlenker and Roberts (2009) and PRISM (County level)

[d]Reproduced from Palmer Z Index of NOAA (State level)

Table 4 Summary statistics for the full sample

Variable	Mean	Std. Dev.	Min.	Max.
Dependent variable				
SCLProducerCount	0.70	2.20	0.00	38.00
Insurance data				
OptionalUnit	0.44	0.25	0.00	1.00
RevenueInsurance	0.64	0.29	0.00	1.00
CoverageLevel	0.67	0.07	0.50	0.85
AcresInsured	92.14	110.19	0.00	1740.84
Profit data				
NetIncome	− 0.19	0.43	− 6.42	13.74
Yearly weather data				
Dday30C	16.00	17.98	0.00	170.23
Drought	0.73	0.61	0.00	3.62
Wetness	1.02	0.78	0.00	3.78
Other Monthly weather data				
Apr_prec	0.89	0.55	0.01	5.85
May_prec	1.04	0.61	0.00	5.55
Jun_prec	1.07	0.61	0.00	7.28
Jul_prec	0.95	0.56	0.00	4.52
Aug_prec	0.90	0.56	0.00	5.01
Sep_prec	0.85	0.61	0.01	5.36
Oct_prec	0.79	0.56	0.01	5.53
Apr_tMin	5.57	4.42	− 8.38	20.94
May_tMin	10.90	4.05	− 1.42	23.53
Jun_tMin	16.07	3.62	3.34	24.98
Jul_tMin	18.17	3.16	6.80	28.68
Aug_tMin	17.33	3.48	5.17	27.65
Sep_tMin	13.19	3.92	0.46	25.09
Oct_tMin	6.70	4.12	− 5.48	21.56
Apr_tMax	19.29	4.76	1.54	34.33
May_tMax	24.04	3.88	11.38	37.88
Jun_tMax	28.78	3.43	17.88	42.16
Jul_tMax	30.95	2.94	21.08	43.66
Aug_tMax	30.34	3.17	20.24	43.59
Sep_tMax	26.71	3.39	16.39	40.52
Oct_tMax	19.88	4.57	5.13	36.18

All variables have 29,799 observations from 2162 counties and years 2002–2015 where X_{it-1} is a vector of explanatory variables including the year fixed effects.[24] The term α_i is the county fixed effect to control for the time-invariant county-level unobserved heterogeneity. For example, similar claim filing practices may be common in a county, since producers' attitudes can be influenced by peers and neighbours within the same county.[25] Finally, ε_{it} denotes an idiosyncratic error term. The linear model does not take into account the count data generating process and the fact that the dependent variable is left-censored. However, it requires the least assumptions among the three models we estimate.

The second model we estimate is the fixed effects Poisson model, which accommodates the first two features of our data discussed above. More specifically, in this model we assume y_{it} follows the Poisson probability density distribution,

$$f\left(y_{it}|\lambda_{it-1},\alpha_i\right)=\frac{\exp\left(-\alpha_i\lambda_{it-1}\right)\cdot\left(\alpha_i\lambda_{it-1}\right)^{y_{it}}}{y_{it}!},$$

where $\lambda_{it-1}=\exp\left(X'_{it-1}\Theta\right)$. A nice feature of the Poisson model is that Wooldridge (1999) shows that estimates from the Poisson maximum likelihood estimation (MLE) are consistent as long as the conditional expectation assumption $E\left(y_{it}|X_{it-1},\alpha_i\right)=\alpha_i\exp\left(X'_{it-1}\Theta\right)$ holds.

As a result, the fixed effects Poisson model requires fewer assumptions than other alternatives. In contrast, another popular count data model, the fixed effects negative binomial (NB) model (Hausman et al. 1984), requires the full distributional assumption for the estimates to be consistent.

Finally, to accommodate all three data features at the same time, we employ the random effects Poisson model with censoring. In this model, the probability function for the observed dependent variable y_{it} (i.e., the censored number of producers on the SCL) is

$$\Pr\left(y_{it}|X_{it-1};\Theta,\alpha_i\right)=\begin{cases}f\left(y^*_{it}=0,1\ldots,\text{or}L|X_{it-1};\Theta,\alpha_i\right) & \text{if}\,y_{it}\le L,\\ f\left(y^*_{it}=y_{it}|X_{it-1};\Theta,\alpha_i\right) & \text{if}\,y_{it}>L,\end{cases}\tag{2}$$

where y^*_{it} is the latent dependent variable (i.e., the true number of producers on the SCL) left-censored at L. In this case,

$$\begin{aligned}\Pr\left(y_{it}\le L|X_{it-1};\Theta,\alpha_i\right)&=f\left(0|X_{it-1};\Theta,\alpha_i\right)+f\left(1|X_{it-1};\Theta,\alpha_i\right)\\&+\cdots+f\left(L|X_{it-1};\Theta,\alpha_i\right)=g\left(L|X_{it-1};\Theta,\alpha_i\right).\end{aligned}$$

Therefore, assuming the independence between observations conditional on covariates and unobserved heterogeneity, the likelihood function for all observations from county i is

[24] Note that the explanatory variables (X) pertains to year $t-1$ since the SCL for year t is determined using data through December of year $t-1$. For example, the SCL in year $t=2018$ is finalised some time in April 2018 and is determined by using data through December 2017.

[25] For example, an "every one does it" attitude may result in a higher potential risk of fraud, waste, and abuse.

$$\Pr\left(y_{i1}\ldots,y_{iTi}|X_{i1},\ldots X_{i(T-1)_i}\Theta,\alpha_i\right)=\prod_{\substack{t=1\\t=1}}^{T_i}f\left(y_{it}|X_{it-1},\Theta,\alpha_i\right)^{d_{it}}\cdot g\left(L|X_{it-1};\Theta,\alpha_i\right)^{(1-d_{it})}$$
$$=\prod_{t=1}^{T_i}\left[\frac{\exp\left(-\alpha_i\lambda_{it-1}\right)\left(\alpha_i\lambda_{it-1}\right)^{y_{it}}}{y_{it}!}\right]^{d_{it}}\cdot g\left(L|X_{it-1};\Theta,\alpha_i\right)^{(1-d_{it})},$$

where $d_{it}=1$ if $y_{it}>L$, $d_{it}=0$, otherwise.

Assuming that the unobserved heterogeneity α_i follows a distribution $h(\alpha_i)$, we can further derive the likelihood function conditional only on the observed covariates as,

$$\Pr\left(y_{i1},\ldots,y_{iTi}|X_{i1},\ldots,X_{i(T-1)i};\Theta\right)$$
$$=\int\prod_{t=1}^{T_i}\left[\frac{\exp\left(-\alpha_i\lambda_{it-1}\right)\left(\alpha_i\lambda_{it-1}\right)^{y_{it}}}{y_{it}!}\right]^{d_{it}}\cdot g\left(L|X_{it-1};\Theta,\alpha_i\right)^{(1-d_{it})}h\left(\alpha_i\right)d\alpha_i \tag{3}$$

In order to approximate the integral in (3), we further assume that the unobserved heterogeneity α_i follows a log-normal distribution with log mean μ and log variance $\sigma^2\left(\text{i.e., } \alpha_i^m=\exp\left(\mu+\sigma u_i^m\right)\right)$ and draws $u_i^m, m=1,2,\ldots,M$ randomly from a standard normal distribution based on Halton sequences.[26] Then we maximise the following log-likelihood function,

$$\sum_{i=1}^{N}\log\left\{\frac{1}{M}\sum_{m=1}^{M}\prod_{t=1}^{T_i}\left[\frac{\exp\left(-\alpha_i^m\lambda_{it-1}\right)\left(\alpha_i^m\lambda_{it-1}\right)^{y_{it}}}{y_{it}!}\right]^{d_{it}}\cdot g\left(L|X_{it-1};\Theta,\alpha_i^m\right)^{(1-d_{it})}\right\},$$

where N is the number of counties in our data set.

Furthermore, there could be both spatial and temporal correlation among the residuals that need to be accounted for. To account for the potential bias in standard errors from these dependencies, we employ an (overlapping) block bootstrapping procedure based on contiguous counties.[27] To do this, we randomly choose a county with replacement and then select all yearly observations from this county and all of its contiguous counties.[28]

[26] For further discussion on Halton draws, see Cappellari and Jenkins (2006), Haan and Uhlendorff (2006), and Train (2003).

[27] This allows for dependence within a contiguous group but assumes non-contiguous counties are independent or weakly dependent. For an introductory survey on block bootstrap approaches, see Hall et al. (1995).

[28] We will also examine the sensitivity of our results using alternative clustering procedures (e.g. one-way/two-way clustering, spatial heteroscedasticity and autocorrelation consistent (HAC) approach, and block bootstrap at the state level) and discuss them in the robustness check section below.

Endogeneity

The second challenge we face in our empirical analysis is that several of the explanatory variables are potentially endogenous. First, the county-level percentage of acres insured as optional units is likely to be endogenous, since there may be unobservable latent variables that influence both this variable and the dependent variable y_{it} (i.e., the number producers on the SCL). For example, if on-site inspections were heavily enforced in the preceding year, but differently among acres insured under unit structures (e.g., focused on acres insured as optional units), then in the current year, producers may be less likely to insure acres as optional units, and the number of producers on the SCL may also be affected. Similarly, average coverage level, percentage of acres insured under revenue-based policies, and number of acres insured are also likely to be endogenous as they are all decision variables by the producers. In addition, the net income variable depends on farmers' decisions for input expenditure and hence is also potentially endogenous.

To correct for the potential bias caused by endogeneity, for the linear panel data fixed effects model we use the two-stage least squares (2SLS) estimator. For the other two non-linear models we employ the control function approach (Hausman 1978; Wooldridge 1997, 2014; Terza et al. 2008). More specifically, the first stage of the control function approach is the same as that of 2SLS where each endogenous variable is regressed on the exogenous variables in the model and a set of instrumental variables. After estimation, the residual is retained. In the second stage of the control function approach, the residuals from all first-stage regressions for the endogenous variables are included in the main regression as additional explanatory variables. Terza et al. (2008) show that in a non-linear framework the control function estimator is consistent, while the non-linear 2SLS approach is not. The intuition of the control function approach is that it divides the variation in each endogenous variable into two parts. The first one is the portion of the variation explained by the set of exogenous and instrumental variables. The second one is the remaining variation and source of endogeneity. By including the residual from the first-stage regression in the main model, the remaining variation in the endogenous variable can be regarded as exogenous. The bootstrapping approach also corrects for the bias in standard errors from the second-stage estimations in both the 2SLS and the control function approaches.[29]

To implement the control function approach, we created several instrumental variables. USDA-RMA's publicly available county-level crop insurance experience data (i.e., the Summary of Business) report the amounts of subsidised premiums for policies with different characteristics. We divided all the policies in the data into eight groups along the following three dimensions: optional versus non-optional units, low (below 70%) versus high (from 70 to 85%) coverage levels, and revenue-based versus yield-based policies. We then created eight per-acre subsidy variables for the eight groups, one for each group. But in the estimation below, we only use six (first-lagged) variables because two combinations (the combination of optional units, low

[29] For details, see Guan (2003).

coverage level, and revenue-based policies, and the combination of optional units, high coverage level, and yield-based policies) have relatively smaller numbers of observations and using them would result in a significant loss of observations. Due partly to the fact that insurance premiums in a county are determined by the loss cost ratio (i.e., indemnities/liabilities) history of the county and subsidy amounts are percentages of the premiums, there are both temporal and county-level variations in the per-acre subsidy variables. Table 5 presents the summary statistics for the per-acre subsidy variables. The amounts of subsidies for different kinds of policies influence which type of policies producers purchase and hence the county-level percentage of acres insured as optional units, the average coverage level and the percentage of acres insured under revenue rather than yield-based policies. Also, the overall increase and decrease of the subsidies affect how many acres are insured. On the other hand, it is unlikely for the subsidies to have a direct effect on the number of producers on the SCL.

The second set of instruments we use is the lagged weather variables. The weather instruments include both the linear and quadratic precipitation and temperature variables. They are expected to capture the second-order effects of weather conditions on the endogenous variables above. We also note that when farmers choose insurance policies and make decisions on inputs in the current year, their decisions may be influenced by their experiences in the previous year, especially the weather conditions. On the other hand, it is unlikely that weather conditions of the past year would affect producers' claims filing behaviour and hence the number of producers on the SCL in the current year.

Results

We first estimated the three econometric models, assuming that all explanatory variables are exogenous. The coefficient estimates are reported in Table 6. The regressions for the fixed effects linear and Poisson models include all variables listed in Table 4 except for the monthly weather data.[30] The year dummies are also included, but the coefficients for year dummies are omitted for brevity. On the other hand, to avoid computational difficulties, the regression for simulated maximum likelihood estimation (MLE) does not include year dummies. The full results are reported in Table A.1 in the Online Appendix. In addition, to make the results from the three models comparable with one another, using the estimates, we further computed the semi-elasticity of the dependent variable with respect to each explanatory variable, that is, $\frac{\partial \ln \mathbb{E}\left(y_{it} \middle| \cdot\right)}{\partial x_{it-1}}$, where x_{it-1} is one explanatory variable. For the linear fixed effects model, the semi-elasticities and the coefficient estimates are different, while the two are the same for the other two models. The semi-elasticity has the interpretation as the effect of one unit increase in x_{it} on the percentage change in $\mathrm{E}(y_{it})$. All standard

[30] Specifications with only the monthly weather variables or some of the monthly and yearly weather variables are explored in the robustness check section.

Table 5 Summary statistics for the per-acre subsidy variables from 2002 to 2015

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Optional + Low Level + Revenue	24,701	18.053	10.972	0.562	174.050
Optional + Low Level + Yield	26,113	15.184	17.480	0.832	312.778
Optional + High Level + Revenue	25,996	25.836	13.463	4.227	184.789
Optional + High Level + Yield	23,218	25.233	40.221	2.562	1211.684
NonOptional + Low Level + Revenue	26,429	15.765	9.750	1.199	108.333
NonOptional + Low Level + Yield	28,912	9.427	19.937	0.338	901.853
NonOptional + High Level + Revenue	27,133	25.591	14.410	2.867	169.497
NonOptional + High Level + Yield	25,397	24.642	46.502	1.857	1311.519

errors are obtained by the (overlapping) block bootstrap procedure based on the contiguous counties as discussed in the section "Estimation".

As we can see from the table, the results are, in general, robust across different models in terms of the signs of the coefficient estimates. However, the magnitudes and statistical significance of the estimates do differ somewhat across different models. As the simulated ML estimator takes into account all three of the main features of our data set, we focus our discussion on results from this model. First, a 1% increase in the percentage of acres insured as optional units increases the number of producers on the SCL by 0.62%. Second, when the average coverage level in a county increases by 1%, the number of producers on the SCL in the county will increase by 5.12%. Third, the percentage of acres insured under the revenue-based rather than yield-based policies is estimated to have a negative and statistically significant effect on the number of producers on the SCL. Fourth, a 1000 increase in the number of acres insured increases the number of producers on the SCL by 0.35%. This result implies that the incidence of anomalous claims behaviour increases with the number of acres insured.

With regard to other factors, our results show that net income has a negative effect on anomalous claims behaviour. When the net income for the county increases by USD 100 million, the number of producers on the SCL decreases by 15.7%. Our results also reveal that extreme weather events in a particular period increase anomalous claims behaviour. Specifically, the more severe or extreme weather that producers encounter during April-October, the higher the number of producers on the SCL. Most of these extreme weather variables are statistically significant at the 1% or 5% level. These findings provide strong evidence that unfavourable weather conditions lead to more anomalous claims behaviour. Insured producers seem to be taking advantage of these adverse weather conditions as the backdrop to submitting anomalous claims (since it is harder for non-legitimate claims to be identified in this situation).

Table 6 Estimation results ignoring endogeneity

	Dependent variable: number of SCL producers			
	FE Linear		FE Poisson	Simulated MLE
	Coefficient	Semi-elasticity		
Insurance data				
OptionalUnit	0.186	0.267	0.951***	0.618***
	(0.199)	(0.285)	(0.311)	(0.101)
CoverageLevel	5.00***	7.178***	12.929***	5.115***
	(1.376)	(1.975)	(2.093)	(0.403)
RevenueInsurance	−0.252	−0.361	−0.286	−0.635***
	(0.267)	(0.384)	(0.354)	(0.127)
ln(AcresInsured)	0.231***	0.332***	1.571***	0.348***
	(0.84)	(0.121)	(0.252)	(0.024)
Profit data				
NetIncome	−0.317	−0.455	−0.328***	−0.157***
	(0.201)	(0.289)	(0.144)	(0.048)
Yearly weather data				
Dday30C	0.007***	0.010***	0.011***	0.003***
	(0.003)	(0.004)	(0.003)	(0.001)
Drought	0.163***	0.233***	0.212***	0.062**
	(0.057)	(0.081)	(0.058)	(0.030)
Wetness	0.071*	0.101*	0.069	0.033
	(0.038)	(0.055)	(0.046)	(0.021)
Obs.	29,799	29,799	15,145	29,799
Counties	2162	2162	1083	2162

Parentheses: standard errors block bootstrapped based on the contiguous counties from 100 bootstrap samples (25 bootstrap samples in the case of simulated MLE)

Parameter estimates for year dummies and constant are omitted for the sake of brevity. See the Online Appendix Table A.1 for full estimation results

FE fixed effects

***p <0.01, **p <0.05, *p <0.10

Correcting for endogeneity

We then re-estimated the three models above using the 2SLS or the control function approach to correct for the possible bias from the potentially endogenous variables. In the first stage of the 2SLS or the control function approach, each endogenous variable is regressed on the instrumental variables (the six lagged subsidy variables and lagged weather variables in our context) and the exogenous variables in our main model. Results from the first-stage regressions are provided in Table A.2 in the Online Appendix. As we can see from this appendix table, by and large, the instrumental variables appear to be associated with the potentially endogenous variables with the expected signs and they are statistically significant. For example, higher per-acre subsidy for the optional units leads to an increase in the percentage of acres insured as optional units. On the contrary, per-acre subsidy for non-optional units shows the opposite effect on the percentage of acres insured as optional units. Table A.2 also reports the F statistics for the null hypothesis that the coefficients of the instrumental variables are jointly equal to zero. All of these first-stage F statistics have very small p-values and we can reject the null hypothesis that our instrumental variables are weak.

Results from the second stage of 2SLS or the control function approach are reported in Table 7 with the associated semi-elasticities for the fixed effects linear model.[31] Again, the results are robust across different models in terms of the signs of the coefficient estimates. Therefore, we again focus our discussion on the results from the simulated ML estimator reported in Table 7. Compared with the results in Table 6, we note that our main results are robust with respect to whether endogeneity is controlled for or not in terms of the signs and statistical significance of the coefficient estimates. The coefficient magnitudes, however, do somewhat differ for all variables.[32] For example, a 1% increase in the average coverage level, and a 1000 increase in the number of acres insured, increase the number of producers on the SCL by 6.01% and 0.39%, respectively. A 1% increase in the proportion of optional unit acres also increases the number of SCL producers in the county by 4.45%. On the other hand, a 1% increase in the percentage of acres insured with revenue-based policies (rather than yield-based policies) decreases the number of producers on the SCL by 1.24%. In addition, the results in Table 7 show a larger effect of county-level net income on anomalous claims behaviour. It is estimated that when the net income for the county increases by USD 100 million, the number of producers on the SCL decreases by 19.1%. These results show that it may be important to correct for bias from potential endogeneity when estimating the effects of insurance contract variables on anomalous claims behaviour. Finally, with regard to the effects of adverse weather, our results from Table 7 again show that extreme weather events in a particular period are associated with more anomalous claims behaviour.

Robustness checks

First, we examine the sensitivity of our model results when alternative clustering procedures are used to calculate standard errors. In this robustness check we still use the same empirical specification used in the main results section, except that we employ different clustering procedures for calculating the standard errors. In particular, for the FE linear model that ignores endogeneity we report spatial HAC standard errors that account for both spatial and serial dependence, as developed in

[31] The full results for Table 7 are reported in Table A.3 in the Online Appendix.

[32] The FE linear and Poisson models generally produce larger coefficients when potential endogeneity is controlled for, compared to when potential endogeneity is ignored. But note that the magnitudes of the coefficient estimates from the simulated MLE model are fairly similar when endogeneity is accounted for and when it is not.

Table 7 Estimation results correcting for endogeneity

	Dependent variable: Number of SCL producers			
	FE linear (2SLS)		FE Poisson	Simulated MLE
	Coefficient	Semi-elasticity	(CF)	(CF)
Insurance data				
OptionalUnit	1.416	1.524	1.770*	0.445**
	(1.223)	(1.316)	(0.928)	(0.182)
CoverageLevel	25.484***	27.431***	23.166***	6.006***
	(6.648)	(7.156)	(6.308)	(0.462)
RevenueInsurance	− 1.838	− 1.978	− 1.032	− 1.235***
	(1.613)	(1.736)	(1.454)	(0.214)
ln(AcresInsured)	1.010	1.088	2.533**	0.390***
	(1.076)	(1.159)	(1.030)	(0.034)
Profit data				
NetIncome	− 1.270**	− 1.367**	− 0.764	− 0.191**
	(0.638)	(0.687)	(0.690)	(0.085)
Yearly weather data				
Dday30C	0.007	0.007	0.011**	0.006***
	(0.005)	(0.006)	(0.006)	(0.001)
Drought	0.229***	0.247***	0.210***	0.046
	(0.069)	(0.074)	(0.063)	(0.032)
Wetness	0.061	0.065	0.053	0.027
	(0.049)	(0.053)	(0.048)	(0.022)
Obs.	19,387	19,387	11,793	19,387
Counties	1884	1884	970	1884

Parentheses: standard errors block bootstrapped based on the contiguous counties from 100 bootstrap samples (25 bootstrap samples in the case of simulated MLE)

Parameter estimates for residuals from the first-stage estimation, year dummies, and constant are omitted for the sake of brevity. See the Online Appendix Table A.3 for full estimation results

CF control function approach

***p <0.01, **p <0.05, *p <0.10

Conley (1999, 2008), Hsiang (2010), and Fetzer (2014). In this case, we use a distance threshold of 250/500 km[33] and a 14-year time lag. We also report more conservative standard errors by clustering at both the county and state-year level. For the FE Poisson and simulated MLE models, we present block bootstrapped standard errors at the state level.[34] Using block bootstrapping at the state level generally leads

[33] As discussed in Fetzer (2014), it is assumed that spatial correlation is linearly decreasing in the distance from county centroids up to a threshold distance (e.g. 250 or 500 km).

[34] We also examined a block bootstrapped approach at the "state-year" level (e.g. 2-year length, 3-year length, or 5-year length). The implications from those results are similar to the block bootstrapped models at the state level. These results are available from the authors upon request.

to less statistically significant coefficient estimates compared to other less conservative clustering procedures. In our case, this is mainly because the bootstrapping procedure is only based on about 45 blocks (one state being a block).

To summarise, in Table A.4 in the Online Appendix we present robustness check results for the case where endogeneity is ignored and where standard errors are calculated based on the following alternative clustering procedures: (i) FE linear model using the one-way clustering at the county, two-way clustering at the county and state-year level, and the spatial HAC approach; (ii) FE (fixed effects) Poisson model using the simple bootstrap at the county level and block bootstrap at the state level; and (iii) simulated MLE using the simple bootstrap at the county level and block bootstrap at the state level. In Table A.5 in the Online Appendix we provide the robustness check results when endogeneity is addressed using 2SLS and control function approaches and using all the alternative clustering procedures enumerated above (with the exception of the spatial HAC).

Not surprisingly, we find that using a smaller number of clusters when calculating standard errors leads to larger standard error magnitudes and consequently less statistical significance in the key variables (See Tables A.4 and A.5 in the Online Appendix). Nonetheless, we note that unit structure, coverage level, acres insured, net income, extreme heat (i.e., Dday30C), and drought variables are consistently statistically significant across these alternatives. In particular for the preferred simulated MLE models, the key insurance variables are strongly statistically significant at the 1% significance level, regardless of the approach to inference.

Moreover, in Table A.4 (second column) in the Online Appendix we also explored the sensitivity of our results when including an interaction term between coverage level and unit structure.[35] For the FE linear model using one-way clustered standard errors, the interaction term was not statistically significant even though this standard error calculation procedure tends to already produce the smallest standard errors (and hence more statistical significance). Although not presented here, this interaction term is also statistically insignificant when using all the other estimation and standard error calculation procedures. This is the main reason why this interaction term was not included in the main specification above.

Second, we conduct robustness checks to examine the sensitivity of our results when two alternative weather specifications are used in the empirical model: (i) only using monthly temperature and precipitation variables (i.e., instead of the 3 yearly weather variables in the main specification);[36] and (ii) a combination of the yearly and monthly extreme weather variables.[37] In this robustness check we only present the results from the FE linear and Poisson models, because the simulated MLE suffers from convergence issues when we use the monthly weather variables. Tables

[35] A referee suggested this specification, since crop insurance subsidy rates may vary by both coverage level and unit structure. We are grateful for this suggestion.

[36] In these empirical specifications we include quadratic precipitation and temperature terms, in addition to the linear terms of these variables.

[37] We use the Drought/Wetness variables from yearly weather conditions and tMin/tMax variables from the set of monthly weather measures. In this specification we only consider the linear temperature terms.

A.6 and A.7 in the Online Appendix report the main results from each alternative specification of the weather variables. We note that in terms of the signs and statistical significance of the coefficient estimates our main findings are robust to the different empirical weather specifications explored.

Lastly, Table A.8 in the Online Appendix presents the robustness check results when using alternative coverage level cut-offs in constructing the subsidy instruments used in the procedures that address endogeneity. In the main specification we used a 70% coverage level cut-off for the subsidy instruments (i.e., “low” for below 70%, and “high” for 70% to 85%). In this robustness check we consider three alternative cut-offs: (i) 65%, (ii) 75%, and (iii) the median coverage level of the county. We find that the signs of the coefficient estimates are still consistent with those in the main specification (See Online Appendix Table A.8). On the other hand, due mainly to the fact that more observations are dropped in these estimations compared to the case with 70% coverage level used as the cut-off, there are generally fewer statistically significant coefficient estimates in these cut-off robustness checks compared to the main results above.[38] Nevertheless, the key contract element and growing condition variables that are statistically significant in the main results section are still mostly significant under alternative subsidy cut-off levels.

Conclusions

This study empirically examines whether certain crop insurance contract elements and growing conditions influence the occurrence of anomalous claims behaviour in the U.S. crop insurance programme. “Anomalous claims behaviour” is measured here using the number of producers included in the SCL for a county. These SCL producers are identified based on claims behaviour that is not consistent with other producers in a county who produce the same crop and use the same practices (i.e., hence they are deemed “anomalous”). County-level variables representing contract elements and growing conditions are merged together with the SCL data to construct a county-level panel data set that is used to achieve the study objective. Several econometric procedures are then implemented to accommodate several unique features of this county-level data set (e.g. the count nature and censoring in the dependent SCL variable, potential endogeneity of the independent variables, and spatial/temporal correlations), which consequently allows more precise parameter estimates.

Our results provide strong evidence that several contract components chosen by producers strongly affect the extent of anomalous claims behaviour in a county. For instance, as the percentage of acres insured as optional units increases, the number of SCL producers in a county also increases. Our analyses also show that average

[38] For example, the number of observations in the case of 70% cut-off in the FE linear model (or FE Poisson) is 19,387 from 1884 counties (11,793 from 970 counties). However, it is reduced to 14,996 from 1593 counties (9511 from 850 counties) in the case of using county median-based cut-off. This is because quite a few observations are at the median level.

county-level coverage levels and the total number of insured acres in a county are also positively related to the number of SCL producers (and consequently the degree of anomalous claims behaviour in a county). These contract elements strongly related to anomalous claims behaviour provide indication of potential underwriting vulnerabilities that may merit further investigation and study. The adverse claims behaviour that is likely due to optional unit choice suggests that the premium “surcharge” for insuring optional units may not be adequate. There is a need for an updated analysis akin to the approach in Knight et al. (2010) but utilising more recent data that account for the higher enterprise unit subsidies implemented in the 2008 Farm Bill. Re-examining the coverage-level relativities used in establishing premiums for yield- and revenue-based policies may also be appropriate given the finding that coverage level choices have an impact on anomalous claims behaviour. Moreover, these potential underwriting vulnerabilities may also be utilised to improve the compliance efforts of RMA. Producers already in the Spot Check List with high coverage levels (say 80–85%), many optional units, yield-based policies, and large acreage may automatically be flagged such that on-site inspections for these insureds will be required.

The growing condition results, where anomalous claims behaviour tends to increase during general downturns in the agricultural economy and during extreme weather events, imply that use of more compliance resources during these periods would also be appropriate. More on-site and in-season inspections for SCL producers during economic downturns and in areas with widespread adverse weather events may be a good use of compliance resources. In addition, it may also be reasonable to “cast a wider net” in terms of the number of producers to include in the SCL for years with economic downturns and areas with widespread adverse weather, given that our findings suggest that insureds may be taking advantage of these adverse conditions to possibly misrepresent claims.

Finally, although our research has provided an initial first step in understanding factors that are correlated with anomalous claims behaviour (and potential fraud, waste, and abuse) in crop insurance, there are still a number of directions for future research. An analysis using individual-level administrative data would be a potentially fruitful avenue for future research. In addition, a crop-specific evaluation of factors that influence anomalous claims behaviour (using individual or county-level data) would be useful. Third, given the recommendations above for increased inspections and auditing, another topic for future research is to examine how RMA inspections and auditing strategies actually affect anomalous claims behaviour.

Acknowledgements This work was supported by USDA-NIFA Grant No. 2016-67023024618. Rejesus was also supported by USDA Hatch Project No. NC02696 and Multi-State Hatch Project No. NC01177. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the U.S. Department of Agriculture.

References

Babcock, B.A., and C.E. Hart. 2005. ARPA subsidies, unit structure, and reform of the U.S. crop insurance program. Briefing Paper 05-BP 45, Center for Agricultural and Rural Development, Iowa State University, Ames, IA.

Cappellari, L., and S.P. Jenkins. 2006. Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *The Stata Journal* 6 (2): 156–189.

Coble, K.H., T.O. Knight, R.D. Pope, and J.R. Williams. 1997. An expected-indemnity approach to the measurement of moral hazard in crop insurance. *American Journal of Agricultural Economics* 79 (1): 216–226.

Conley, T.G. 1999. GMM estimation with cross sectional dependence. *Journal of Econometrics* 92 (1): 1–45.

Conley, T.G. 2008. Spatial econometrics. In *The new Palgrave dictionary of economics*, ed. S.N. Durlauf and L.E. Blume, 741–747. New York: Palgrave MacMillan.

Fetzer, T. 2014. Can workfare programs moderate violence? Evidence from India. Economic Organisation and Public Policy Discussion Papers, EOPP 53.

Guan, W. 2003. From the help desk: Bootstrapped standard errors. *The Stata Journal* 3 (1): 71–80.

Haan, P., and A. Uhlendorff. 2006. Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood. *The Stata Journal* 6 (2): 229–245.

Hall, P., J.L. Horowitz, and B. Jing. 1995. On blocking rules for the bootstrap with dependent data. *Biometrika* 82 (3): 561–574.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46 (6): 1251–1271.

Hausman, J.A., B.H. Hall, and Z. Griliches. 1984. Econometric models for count data with an application to the patents-R&D Relationship. *Econometrica* 52 (4): 909–938.

Hsiang, S.M. 2010. Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences of the United States of America* 107 (35): 15367–15372.

Knight, T.O., and K.H. Coble. 1999. Actuarial effects of unit structure in the U.S. Actual Production History Crop Insurance Program. *Journal of Agricultural and Applied Economics* 31 (3): 519–535.

Knight, T.O., K.H. Coble, B.K. Goodwin, R.M. Rejesus, and S. Seo. 2010. Developing variable unit-structure premium rate differentials in crop insurance. *American Journal of Agricultural Economics* 92 (1): 141–151.

Motamed, M., A. Hungerford, S. Rosch, E. O'Donoghue, M. MacLachlan, G. Astill, J. Cessna, J. Cooper. 2018. Federal risk management tools for agricultural producers: An overview. Economic Research Report No. 250, Economic Research Service (ERS), USDA, Washington, DC.

NOAA (National Oceanic and Atmospheric Administration). 2019. Palmer Drought Severity Index. https://www7.ncdc.noaa.gov/CDO/CDODivisionalSelect.jsp. Accessed 2 Aug 2019.

Roberts, M.J., N. Key, and E. O'Donoghue. 2006. Estimating the extent of moral hazard in crop insurance using administrative data. *Review of Agricultural Economics* 28 (3): 381–390.

Schlenker, W., and M.J. Roberts. 2006. Nonlinear effects of weather on corn yields. *Review of Agricultural Economics* 28 (3): 391–398.

Schlenker, W., and M.J. Roberts. 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences* 106 (37): 15594–15598.

Schnitkey, G., and B. Sherrick. 2014. Coverage levels on crop insurance and the SCO alternative. Farmdoc Daily (4):78, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, April 29, 2014.

Terza, J.V., A. Basu, and P.J. Rathouz. 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics* 27 (3): 531–543.

Train, K. 2003. *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

U.S. Department of Commerce-BEA (Bureau of Economic Analysis). 2019. Regional Economic Accounts, CA45 Farm Income and Expenses. https://www.bea.gov/regional. Accessed 2 Aug 2019.

USDA-RMA (United States Department of Agriculture, Risk Management Agency). 2011. Program Compliance and Integrity. Annual Report to Congress, January-December 2006.

USDA Office of the Inspector General (OIG). 2017. RMA's Utilization of Contracted Data Mining Results. Audit Report No. 05601-0005-31, USDA OIG, Washington, DC.
USDA-RMA (Risk Management Agency). 2015. Crop Insurance Handbook.
USDA-RMA (Risk Management Agency). 2018. Federal Crop Insurance Program Spot Check List Study. https://govtribe.com/opportunity/federal-contract-opportunity/federal-crop-insurance-program-spot-check-list-study-rmaslc010418. Accessed 2 Aug 2019.
USDA-RMA (Risk Management Agency). 2019. State/County/Crop Summary of Business (1989–2015). https://www.rma.usda.gov/data/sob/scc/index.html. Accessed 2 Aug 2019.
Walters, C.G., C.R. Shumway, H.H. Chouinard, and P.R. Wandschneider. 2015. Asymmetric information and profit taking in crop insurance. *Applied Economic Perspectives and Policy* 37 (1): 107–129.
Wooldridge, J.M. 1997. Quasi-likelihood methods for count data. In *Handbook of applied econometrics, Vol 2: Microeconomics*, ed. M.H. Pesaran and P. Schmidt. Malden, MA: Blackwell Publishers Ltd.
Wooldridge, J.M. 1999. Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics* 90 (1): 77–97.
Wooldridge, J.M. 2014. Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics* 182: 226–234.

About the authors

Sungkwol Park is Senior Deputy Director at the Ministry of Economy and Finance in South Korea. He holds a PhD in Economics from North Carolina State University, where he worked on this article and other compliance issues in crop insurance.

Barry K. Goodwin is William Neal Reynolds Distinguished Professor and Graduate Alumni Distinguished Professor in the Departments of Agricultural and Resource Economics and Economics at North Carolina State University. His research interestes include topics in risk management, crop insurance, policy analysis, consumer demand, price analysis, international trade and applied econometrics and statistics. He holds a PhD in Economics from North Carolina State University.

Xiaoyong Zheng is Professor and Director of the Graduate Program in the Departments of Agricultural and Resource Economics and Economics at North Carolina State University. His research interests are in industrial organisation, especially markets with asymmetric information, and applied econometrics. He holds a PhD in Economics from Indiana University.

Roderick M. Rejesus is a Professor and Extension Specialist in the Department of Agricultural and Resource Economics at North Carolina State University. His research interests are in applied production economics, especially topics related to risk management, crop insurance, and economic impact assessment of agricultural technologies and practices. He holds a PhD in Agricultural Economics from the University of Illinois at Urbana-Champaign.